



Giuggioli, L., & Neu, Z. (2019). Fokker-Planck representations of non-Markov Langevin equations: application to delayed systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 377(2153), [2018131].
<https://doi.org/10.1098/rsta.2018.0131>

Peer reviewed version

License (if available):
Other

Link to published version (if available):
[10.1098/rsta.2018.0131](https://doi.org/10.1098/rsta.2018.0131)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the accepted author manuscript (AAM). The final published version (version of record) is available online via the Royal Society at <https://doi.org/10.1098/rsta.2018.0131> . Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Article submitted to journal

Subject Areas:

Statistical physics, Transport processes, Non-Markovianity

Keywords:

Non-Markov processes, Stochastic systems, Delayed Langevin equation, Fokker-Planck equation, Generalised Langevin equation

Author for correspondence:

Luca Giuggioli

e-mail: Luca.Giuggioli@bristol.ac.uk

Fokker-Planck representations of non-Markov Langevin equations: application to delayed systems

Luca Giuggioli^{1,2} and Zohar Neu¹

¹Department of Engineering Mathematics, University of Bristol, Woodland Road, Bristol, BS8 1UB

²Bristol Centre for Complexity Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UB

Noise and time delays, or history-dependent processes, play an integral part of many natural and man-made systems. The resulting interplay between random fluctuations and time non-locality are essential features of the emerging complex dynamics in non-Markov systems. While stochastic differential equations in the form of Langevin equations with additive noise for such systems exist, the corresponding probabilistic formalism is yet to be developed. Here we introduce such a framework via an infinite hierarchy of coupled Fokker-Planck equations for the n -time probability distribution. When the non-Markov Langevin equation is linear, we show how the hierarchy can be truncated at $n = 2$ by converting the time non-local Langevin equation to a time-local one with additive coloured noise. We compare the resulting Fokker-Planck equations to an earlier version, solve them analytically and analyse the temporal features of the probability distributions that would allow to distinguish between Markov and non-Markov features.

1. Introduction

Deterministic tendencies and random fluctuations are ubiquitous and relevant in a variety of contexts from engineering to biology. A vast number of theoretical studies on their interplay exist in the literature since the early studies by Langevin [1] and Smoluchowski [2]. These two celebrated approaches, although technically very different, are equivalent in their description of a system's dynamics. The former employs a stochastic differential equation model, the Langevin equation (LE), and allows to represent individual system trajectories. The latter combines all trajectories simultaneously and treats them as an ensemble, making use of a deterministic partial differential equation (PDE), the so-called Fokker-Planck (FP) equation [3], to represent the time dependence of the probability distribution of the random variables.

As the equivalence between the two descriptions is well known for Markov processes, predictions and analyses of LE and FP have been a workhorse to study the dynamics of random systems over the last 60 years. For non-Markov processes, on the other hand, while there has been a great deal of studies for deterministic systems since the 60s [4], random systems have received little attention in general. Exceptions have been the analyses of non-Markov linear systems, starting from the seminal work on Delayed Langevin Equation (DLE) by K  chler and Mensch in 1992 [5] and the studies in the 70's and 80's by Fox [6–8], Hanggi [9–12] and Sancho and San Miguel [13–15], as well as others [16,17], to find a *bona fide* FP equation for the so-called Generalised Langevin Equation (GLE), a linear time non-local LE with additive noise. Those initial studies brought attention to the probabilistic-Langevin equivalence problem, but to a great extent it dissipated by the late 80's.

Renewed interest in the equivalence problem has surfaced in the early 2000's with the work of Longtin et al. [18,19] and Budini and C  ceres [20–22]: the former on the formulation of a FP equation for a DLE, and the latter on the probabilistic representation for linear non-Markov systems via characteristic functionals, which do not make use of FP equations. In that vein, analyses of the equivalence of a DLE to a delayed random walk model have been presented by Ohira [23,24] since 1995, while various studies by Frank [25–28] focused on finding analytic solutions of the FP equation for the DLE at steady state.

This renewed interest stems, in part, from the technological advances in a variety of biological areas, bringing empirical observations with ever increasing temporal resolution. The improved resolution in the data, combined with the ever present uncertainties or inability to construct first-principle models of biological phenomena, has made apparent the need for a general framework to represent non-Markov stochastic processes. The equivalence between a LE and FP description is at the heart of that development.

Following the approach of Budini and C  ceres, one of the present authors has shown in ref. [29] the importance of constructing joint-probability distribution function (pdf) to represent non-Markov stochastic processes compared to Markov cases. The joint probability distribution is in fact necessary to construct the conditional probability distribution, which is the so-called *bona fide* representation of the random process. Following on from ref. [30], a FP representation for the one-time and two-time pdf has been developed to construct the conditional pdf for a DLE [31].

Despite these recent findings for linear non-Markov systems, the formulation of a FP description of non-linear, non-Markov processes remains an open problem. Our interest here is to review these recent studies and present some new result for (linear) DLE, and at the same time to set the stage for tackling the more general problem of a probabilistic representations of non-linear, non-Markov systems.

The paper is organised as follows. In Sec. 2 we consider the general problem of a delayed non-linear Langevin equation and show the infinite hierarchy of FP equations that can be constructed from it. In Sec. 3 we present the DLE, we review the solution method and the convolutionless transform. This transform converts the time non-local LE to another linear LE, where the non-Markovianity is transferred from the drift to the noise term, which becomes coloured, while the Langevin equation becomes local in time. In Sec. 4 we write and solve a new set of FP equations

for the DLE, and compare these to the set presented earlier in the literature [31]. Sec. 5 is dedicated to review the method of characteristic functionals that allow, for linear time-non local LE, to find any joint-time pdf without solving a FP equation. The formalism from Sec. 5 is then used in Sec. 6 to present ways to identify ageing effects, characteristic of non-Markov processes. Sec. 7 forms the concluding remarks.

2. The fundamental problem

(a) Markov systems

We begin by considering the dynamics of a stochastic variable, $x(t)$ as a function of time t , taken here to be 1D for simplicity. For Markov cases when the dynamics are governed by a deterministic law, $-\gamma g(x)$, and noise perturbs the system by adding random fluctuations, the prescription to construct a PDE for the probability $P(x, t)$ is well known. When the noise is assumed to be Gaussian white noise, the form of the PDE is the celebrated Smoluchowski equation [2], also called Fokker-Planck equation [3]

$$\frac{\partial P(x, t)}{\partial t} = \frac{\partial}{\partial x} \left[\gamma g(x) + D \frac{\partial}{\partial x} \right] P(x, t), \quad (2.1)$$

where D is the diffusion constant. The dynamics of the system is fully described once the initial condition is known. Given that the FP equation (2.1) is linear in probability space, the initially localised solution $P(x, t|x_0, 0)$, obtained when the initial condition is given by a Dirac delta centred at $x = x_0$, $P(x, 0) = \delta(x - x_0)$, can be generalised to $W(x, t) = \int P(z, t|y, 0)P(y, 0)$ for any initial conditions $P(x, 0)$.

The alternative description of the stochastic system is via the Langevin equation

$$\frac{dx(t)}{dt} = -\gamma g(x) + \sigma \xi(t), \quad (2.2)$$

where γ is a rate and the noise possesses zero mean, i.e. $\langle \xi(t) \rangle = 0$, and correlation $\langle \xi(t_1)\xi(t_2) \rangle = \delta(t_1 - t_2)$, and strength $\sigma = \sqrt{2D}$. The symbol $\langle \rangle$ represents an average over noise realisations, the so-called ensemble average. The equivalence between the two descriptions, that is between the Langevin particle description and the probability of the field variable x , is possible through the microscopic definition of the probability distribution function (pdf)

$$P(x, t) = \langle \delta(x - x(t)) \rangle. \quad (2.3)$$

Differentiation of (2.3) with respect to time gives

$$\frac{\partial P(x, t)}{\partial t} = -\frac{\partial}{\partial x} \left\langle \frac{dx(t)}{dt} \bigg|_{x(t)=x} \delta(x - x(t)) \right\rangle = -\frac{\partial}{\partial x} \langle [-\gamma g(x) + \sigma \xi(t)] \delta(x - x(t)) \rangle, \quad (2.4)$$

which allows to replace $x(t)$ with x given the presence of $\delta(x - x(t))$ as a multiplying factor. Functional calculus (see Appendix A) allows to write $\langle \xi(t) \delta(x - x(t)) \rangle = -\sigma \frac{\partial}{\partial x} \langle \delta(x - x(t)) \rangle / 2$ and one then obtains Eq. (2.1).

Conditional probability

Before we discuss the particle and field formalisms for systems governed by time non-local dynamics, it is useful to discuss more broadly the implications of a probability description for a non-Markov versus a Markov system. For that it is relevant to define the so-called *bona fide* probabilistic representation of a stochastic process [8], that is a representation that defines the state of the system at a future time t , given an initial observation at time $t' < t$. The observation time t' may be different from the initial preparation of the system at time $t = 0$. Such quantity is the conditional probability distribution Q .

To fully describe a system, one is interested in finding the conditional probability of the system at n instances in the future given m observations in the past. Knowledge of the system dynamics

is possible by constructing the general $Q_{n|m}$, which is linked to the joint-probability distribution as follows [32]

$$Q_{n|m}(x_{m+n}, t_{m+n}; \dots; x_{m+1}, t_{m+1} | x_1, t_1; \dots; x_m, t_m)_{x_0} = \frac{P_{n+m}(x_{n+m}, t_{n+m}; \dots; x_1, t_1 | x_0, 0)}{P_m(x_m, t_m; \dots; x_1, t_1 | x_0, 0)}, \quad (2.5)$$

where $m \leq n$ and $t_1 < t_2 < \dots < t_{n-1} < t_n < t_{n+1} < \dots < t_{n+m}$, and where the subindex x_0 indicates that we are describing the situation for which the system was prepared localised at x_0 at time $t = 0$. Iterative use of (2.5) gives $Q_{n|m} = Q_{1|n+m-1} Q_{1|n+m-2} \dots Q_{1|m}$, showing that once $Q_{1|s}$ is constructed, which we might call the s -time conditional pdf, the full dynamics can be obtained.

For Markov systems only the most recent observation defines the dynamics in the future, therefore $Q_{1|s} = Q_{1|1}$. For such cases Eq. (2.5) implies that

$$P_{n+1}(x_{n+1}, t_{n+1}; \dots; x_1, t_1) = \overbrace{Q_{1|1}(x_{n+1}, t_{n+1} | x_n, t_n) \dots Q_{1|1}(x_2, t_2 | x_1, t_1)}^{n\text{-times}} P_1(x_1, t_1). \quad (2.6)$$

Since, with x' being any one of the independent variables x_i inside P_{n+1} , we have $P_n = \int P_{n+1} dx'$, that is P_n is the marginal of P_{n+1} , Markov processes are completely characterised once P_2 is known as $Q_{1|1} = P_2/P_1$. This is not the case for non-Markov processes for which it is necessary to find all $Q_{1|s} = P_{s+1}/P_s$ for all s . This fact is in general true except for certain linear non-Markov systems, for example the class of linear Gaussian processes discussed below, for which it is possible to determine P_{s+1} and P_s .

At the level of the Langevin equation, the distinction between Markov and non-Markov systems can be explained pictorially with Fig. 1, where we display two hypothetical systems, a Markov and a non-Markov one, whose state variable x are prepared localised at $x = 5x_0$. At time $t = s$ both systems are observed for the first time at location $x = x_0$ indicated by the spatial coincidence of all sampled trajectories at time $t = s$. The dynamics of interest is the one for $t > s$ and, depending on the system being Markov or not, prediction abilities may differ. The different colouring scheme of the non-Markov and Markov cases for $t < s$ points to the history dependence of the former versus the latter. For a Markov process, the actual trajectory followed to reach $x = x_0$ is irrelevant for the subsequent dynamics. All that matters is the first observation at $x = x_0$. This is not the case for the non-Markov trajectories, where in principle the entire path taken to reach $x = x_0$ from $x = 5x_0$ defines the dynamics at times $t > s$.

In terms of the conditional probability formalism presented above, this pictorial example indicates that for the Markov case we require knowledge of $Q_{1|1}(x, t | x_0, s)$ since $Q_{1|1}(x, t | x_0, s) = Q_{1|m}(x, t | x_0, s; x_1, t_1; \dots, x_{m-1}, t_{m-1})_{5x_0}$, whereas for the non-Markov case one requires knowledge of $Q_{1|m}$ for all values of m to determine all higher order correlation properties of the system.

(b) Non-Markov systems

A simple way to make the Langevin Eq. (2.2) non-Markov is to include a time non-local dependence of the form

$$\dot{x}(t) = -\gamma f(x(t), x(t-\tau)) + \sigma \xi(t), \quad x(0) = x_0, \quad x(t) = \beta(t), \quad \text{for } -\tau \leq t < 0, \quad (2.7)$$

where $f(y, z)$ is a generic smooth function of (y, z) and $\beta(t)$ is the history function that describes the system deterministically for $t < 0$. If one were to proceed to construct a FP equations as for the Markov case one would encounter the difficulty that $x(t)$ cannot be replaced with the field variable x for all times t because the connection is to values of x at all times in the past [30].

While we cannot deduce a FP of Smoluchowski type from (2.7), it is still possible to write a FP that links the various P_n . Starting from the definition

$$P_n(x_n, t_n; x_{n-1}, t_{n-1}; \dots; x_1, t_1) = \left\langle \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle, \quad (2.8)$$

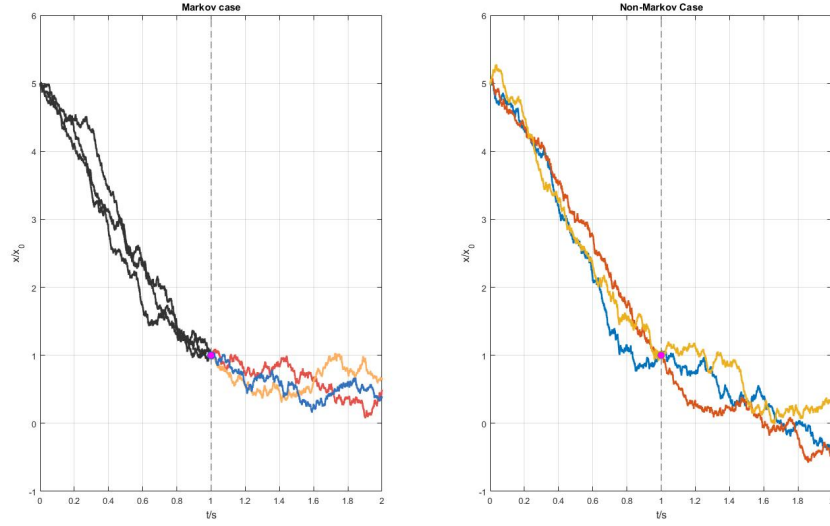


Figure 1. Pictorial comparison of the dynamics of a Markov and a non-Markov system prepared localised at $x = 5x_0$ and observed initially at time $t = s$ at $x = x_0$. For a non-Markov system, in principle, the entire trajectory taken for times $t < s$ may determine which trajectory the system follows for times $t > s$, indicated by the different colours of the paths for time $t > s$. For a Markov system, on the other hand, it is irrelevant. Any of the black paths for $t < s$ may give rise to any of the coloured paths for $t > s$.

the variables x_2, \dots, x_n, \dots represent the system at some other times with $t_{n-1} < t_n$ and differentiating with respect to the latest time t_n one has

$$\begin{aligned} \frac{\partial}{\partial t_n} P_n(x_n, t_n; x_{n-1}, t_{n-1}; \dots; x_1, t_1) = \\ - \frac{\partial}{\partial x_n} \left\langle \left[-\gamma f(x_n, x(t_n - \tau)) + \sigma \xi(t_n) \right] \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle. \end{aligned} \quad (2.9)$$

If one calls $t' = t_n - \tau$ and $x' = x(t_n - \tau)$, it is possible to make the dependence on P_n explicit in the PDE (2.9). The details shown in Appendix A give

$$\begin{aligned} \frac{\partial P_n}{\partial t_n}(x_n, t_n; x_{n-1}, t_{n-1}; \dots; x_1, t_1) = D \frac{\partial^2}{\partial x_n^2} P_n(x_n, t_n; x_{n-1}, t_{n-1}; \dots; x_1, t_1) \\ + \gamma \frac{\partial}{\partial x_n} \int dx' f(x_n, x') P_{n+1}(x_n, t_n; x', t'; x_{n-1}, t_{n-1}; \dots; x_1, t_1). \end{aligned} \quad (2.10)$$

where $t_{n-1} < t' < t_n$ and $x' = x(t')$. Writing out Eq. (2.10) for each n it becomes explicit that one is facing the coupled FP set

$$\begin{aligned} \frac{\partial P_1(x_1, t_1)}{\partial t_1} &= \gamma \frac{\partial}{\partial x_1} \int dx' f(x_1, x') P_2(x_1, t_1; x', t') + D \frac{\partial^2 P_1(x_1, t_1)}{\partial x_1^2}, \\ \frac{\partial P_2(x_2, t_2; x_1, t_1)}{\partial t_2} &= \gamma \frac{\partial}{\partial x_2} \int dx' f(x_2, x') P_3(x_2, t_2; x', t'; x_1, t_1) \\ &\quad + D \frac{\partial^2}{\partial x_2^2} P_2(x_2, t_2; x_1, t_1), \\ &\dots \end{aligned} \quad (2.11)$$

The case $n = 1$ was originally proposed by Guillouzic et al. [18] as a probabilistic representation of the linear delayed Langevin equation with additive Gaussian noise (see below Eq. 3.1).

Despite the simplicity of the FP set (2.12), it is difficult to handle for two reasons. The infinite hierarchy ought to stop at some value of n and conditions to truncate the hierarchy should be developed. At the same time, even with an appropriately chosen truncation, the last FP of the set would depend on one higher moment. The intertwined nature of the joint pdfs in the second term between a lower and higher order P_n in each equation of the set is the result of the non-Markov nature of the dynamics. The n -th truncation, to be of practical use, would thus need to be accompanied with some choice for the quantity $\int dx' f(x_n, x') P_{n+1}(x_n, t_n; x', t'; x_{n-1}, t_{n-1}; \dots; x_1, t_1)$ in terms of lower order P_n 's.

A useful method to truncate the hierarchy consists of manipulating the first term on the r.h.s. of Eq. (2.9). When the function $f(y, z)$ is linear in y and z , this is possible by rewriting the Langevin equation as a weakly non-Markov Langevin equation. In this case the time non-locality is eliminated from the function f , with the noise term becoming correlated. We dedicate the next section to show how this is done in practice.

3. The Green's function for the Langevin dynamics

(a) Example for a linear, time non-local process

As an example of a linear function for $f(y, z)$ we take the Delay Langevin Equation (DLE)

$$\dot{x}(t) = -\gamma x(t - \tau) + \sigma \xi(t), \quad x(0) = x_0, \quad x(t) = \beta(t), \quad \text{for } -\tau \leq t < 0. \quad (3.1)$$

Given the linearity in $x(t)$ of the above equations, it is possible to find the general solution of Eq. (3.1) by considering first the so-called Green's function of the problem, which is simply the solution to the noise-free Langevin equation

$$\dot{\lambda}(t) = -\gamma \lambda(t - \tau), \quad \lambda(0) = 1, \quad \lambda(t) = 0, \quad \text{for } t < 0. \quad (3.2)$$

In other words $\langle x(t) \rangle = \lambda(t)$. The formal solution of the Langevin equation can be written explicitly in terms of the Green's function as

$$x(t) = K(t) + \sigma \mu(t), \quad K(t) = x_0 \lambda(t) + \Psi(t), \quad \mu(t) = \int_0^t \lambda(t - s) \xi(s) ds \quad (3.3)$$

where $\Psi(t) = -\gamma \int_{-\tau}^0 \lambda(t - s - \tau) \beta(s) ds$, that is proportional to the history function (see e.g. [4]). This means that $\Psi(t)$ is identically zero whenever the history function in the DLE is identically zero. Here it should also be noted that the linearity of the system in (3.1) conserves the Gaussian properties of the noise term, $\xi(t)$. Hence, $x(t)$ as defined in (3.3) is itself a Gaussian process.

The regimes of the Langevin equations (3.1) are characterised by the dynamical features of the noise-free equation, that is by the dynamics of the Green's function. There are two stable, and one unstable regimes. In the stable regimes, $\lambda(t)$ is either monotonically decaying to 0, or it displays (damped) oscillation as it decays to 0, while in the unstable regime $\lambda(t)$ displays oscillations with ever increasing magnitude. Examples of the two stable regimes are shown in Fig. 2, where we

have plotted the mean and the mean square displacement (MSD), or variance, for the DLE with a single delay. As the MSD is given by

$$\nu(t) = \left\langle \left(x(t) - \langle x(t) \rangle \right)^2 \right\rangle = \sigma^2 \int_0^t \lambda^2(s) ds, \quad (3.4)$$

the oscillating cases display inflection points in correspondence to when $\lambda(t)$ crosses zero.

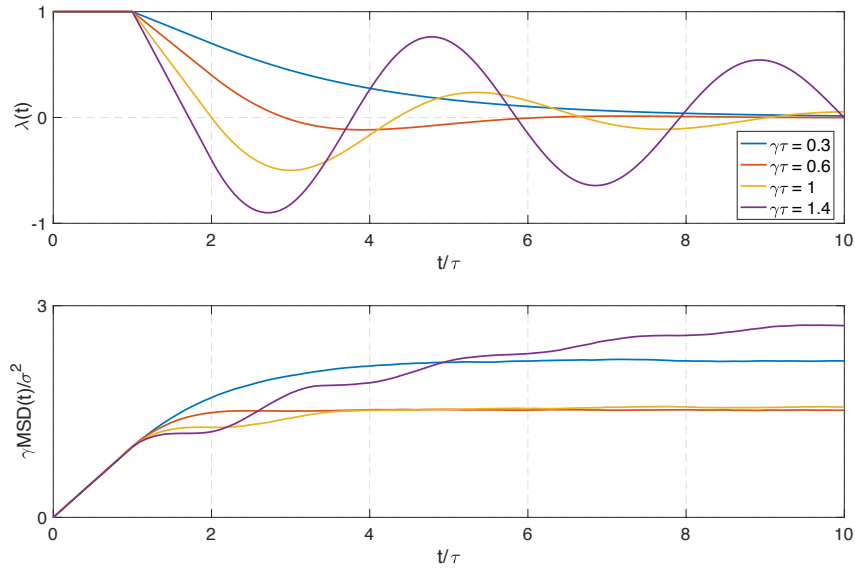


Figure 2. Dynamics of the mean $\langle x(t) \rangle$ and MSD of the DLE (3.1) for different choices of $\gamma\tau$ in the stable regimes. The saturation value of the MSD is known analytically from ref. [5] and equals $\sigma^2[1 + \sin(\gamma\tau)]/[2\gamma \cos(\gamma\tau)]$.

(b) Convolutionless transform

Exploiting the linearity of the solution $x(t)$ in terms of $\lambda(t)$, it is possible to use the so-called convolutionless transform [13] to rewrite the original Langevin equation as a time-local Langevin equation. This consists of differentiating (3.3), thus rewriting (3.1) as

$$\dot{x}(t) = \dot{K}(t) + \sigma \dot{\mu}(t), \quad (3.5)$$

where the noise term $\dot{\mu}(t)$ is no longer white noise, but is instead coloured (see e.g. [33] for an exposition of various forms of coloured noise), with mean and temporal correlation defined by

$$\langle \dot{\mu}(t) \rangle = 0, \quad (3.6)$$

$$\langle \dot{\mu}(t) \dot{\mu}(t') \rangle_{t' < t} = \left\langle \left(\frac{d}{dt} \int_0^t \lambda(t-s) \xi(s) ds \right) \left(\frac{d}{dt'} \int_0^{t'} \lambda(t'-s') \xi(s') ds' \right) \right\rangle. \quad (3.7)$$

The form of Eq. (3.5) is often referred to as a weakly non-Markov Langevin equation since the deterministic part is local in time (the integral over the history, $\beta(t)$, is a known function of time), whereas the non-Markovianity of the stochastic process is hidden inside the noise term.

4. Fokker-Planck equations for the weakly non-Markov Langevin equation

For linear processes such as (3.1), it is possible to use the simplified Langevin form (3.5), so that each element of the FP hierarchy (2.12) is governed by

$$\begin{aligned}\frac{\partial P_n}{\partial t_n} &= -\frac{\partial}{\partial x_n} \left\langle \left[\dot{K}(t_n) + \sigma \dot{\mu}(t_n) \right] \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle. \\ &= -\dot{K}(t_n) \frac{\partial P_n}{\partial x_n} - \frac{\partial}{\partial x_n} \left\langle \left[\sigma \dot{\mu}(t_n) \right] \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle. \quad (4.1)\end{aligned}$$

While the noise term is coloured, making the diffusion term time-dependent, the procedure has succeeded in decoupling the lower and higher order joint pdfs. Each FP in (2.12) can now be solved and any conditional pdf $Q_{1|s}$ can be constructed. We show in detail the FP for P_1 and P_2 , and their analytical solutions.

(a) One and two-time Fokker-Planck equations

Performing the derivation found in Appendix B, we obtain the Fokker-Planck equations for $P_1(x, t|x_0, 0)$ and $P_2(x, t; x', t'|x_0, 0)$, with $t > t'$, given by

$$\begin{aligned}\frac{\partial P_1}{\partial t} &= -\dot{K}(t) \frac{\partial P_1}{\partial x} + \frac{\dot{\nu}(t)}{2} \frac{\partial^2 P_1}{\partial x^2}, \\ \frac{\partial P_2}{\partial t} &= -\dot{K}(t) \frac{\partial P_2}{\partial x} + \frac{\dot{\nu}(t)}{2} \frac{\partial^2 P_2}{\partial x^2} + \frac{\partial \mathcal{C}(t', t)}{\partial t} \frac{\partial^2 P_2}{\partial x \partial x'},\end{aligned} \quad (4.2)$$

where for $t' < t$

$$\mathcal{C}(t', t) = \sigma^2 \int_0^{t'} \lambda(t-s) \lambda(t'-s) ds. \quad (4.3)$$

The FP Eqs. (4.2) are supplemented by the localised initial conditions

$$P_1(x, 0|x_0, 0) = \delta(x - x_0), \quad (4.4)$$

$$P_2(x, t'; x', t'|x_0, 0) = \delta(x - x') P_1(x', t'|x_0, 0), \quad (4.5)$$

which indicates that the solution of the P_1 equation at time t' is needed to generate the initial condition at time $t = t'$ for the P_2 equation.

The simple form of Eq. (4.2) allows to gain physical intuition about the process they describe. In each equation, the first term represents the deterministic, or mean rate of movement for the variable x , which in fact corresponds to the derivative of the mean of the DLE (3.3). The role of this first term, taken as a force resulting from a potential proportional to x , is to drive the solution away from the initial preparation of the system, i.e. with x localised at x_0 , towards $x = 0$. To understand how the dynamics is affected by this first term it suffices to consider the evolution of P_1 , or P_2 , in the absence of diffusion, i.e. $\sigma = 0$, and zero history $\Psi(t) = 0$. Eliminating these terms from the P_1 equation results in the FP solution $P_1(x, t) = \delta(x - x_0 \lambda(t))$, which implies that the system is localised at $x = x_0 \lambda(t)$ for all times $t > 0$. The sign of x_0 determines the direction of the resulting linear force acting on the particle. While $\dot{\lambda}(t) = 0$ for time $t < \tau$, $\dot{\lambda}(t) < 0$ for $t > \tau$. As a result, in the monotonic regime, for $x_0 > 0$ the drift is to the left, while for $x_0 < 0$ it will be to the right. At long times $t \rightarrow \infty$, as $\lambda(t) \rightarrow 0$, this drift term push the solution towards $x = 0$. If $\lambda(t)$ changes sign, the mean of the pdf goes past $x = 0$, but eventually changes direction back towards $x = 0$ once $\dot{\lambda}(t)$ also changes sign. These oscillations continue until the steady state is reached. The same drift dynamics occurs for P_2 .

The second term in both equations represents diffusion in the x -direction. The rate of this diffusion is the rate of change of the variance, or MSD. Since $\dot{\nu}(t) = \sigma^2 \dot{\lambda}^2(t)$ by definition, the diffusion rate decreases to zero as $t \rightarrow \infty$ with the width of the pdf having a finite width. As the

drift term becomes zero, moving the mean of the pdf towards $x = 0$, the steady state of $P_1(x, t)$ is a non-homogeneous pdf centred around $x = 0$.

To interpret the final term in Eq. (4.2), it is helpful to consider the meaning of a mixed partial derivative of a function, f with respect to x and x' . This represents how the slope of f changes along the x direction as one moves along the x' direction (and vice-versa). Equivalently, it can be thought of as a measure of the curvature of the function f in the $x - x'$ plane. Hence, in the dynamic context of Eq. (4.2), the final term is a torque that twists the solution in the $x - x'$ plane. Physically, this term represents the influence that attaining a state x' at time t' has on the probability that the state x is attained at the later time $t > t'$. The strength of this torsion is proportional to the rate of change of the covariance for the process.

The absence of mixed partial derivatives from the usual Fokker-Planck representation of one-dimensional Markov processes in (2.1), suggests that the presence of this term is a consequence of non-Markovianity. Finally it should be noted that since, at long times, the decay of $\lambda(t) \rightarrow 0$ causes $C(t', t)$ to become constant for $t' < \infty$, the coefficient of the mixed derivative term vanishes, and so the twisting of the solution disappears for $t \rightarrow +\infty$.

A peculiarity of Eq. (4.2) — a direct result of the non-Markovianity of the process — is the dependence of the parameters of the FP on the localised initial preparation x_0 . We recall that $\dot{K}(t) = x_0 \dot{\lambda}(t) + \dot{\Psi}(t)$. A way to eliminate the dependence on x_0 from the coefficients of the FP equations exists. It consists of rewriting the ensemble average of Eq. (3.3) as

$$x_0 = \frac{x(t)}{\lambda(t)} - \frac{\Psi(t)}{\lambda(t)} \quad (4.6)$$

and substituting it into (3.5) to obtain a different time-local Langevin equation

$$\dot{x}(t) = -A(t)x(t) + B(t) + \sigma\zeta(t), \quad (4.7)$$

with

$$A(t) = -\frac{\dot{\lambda}(t)}{\lambda(t)}, \quad B(t) = -\frac{\dot{\lambda}(t)}{\lambda(t)}\Psi(t) + \dot{\Psi}(t), \quad (4.8)$$

and the noise term and its correlation now being

$$\begin{aligned} \zeta(t) &= \lambda(t) \frac{d}{dt} \int_0^t ds \frac{\lambda(t-s)}{\lambda(t)} \xi(s) \\ \langle \zeta(t) \rangle &= 0, \quad \langle \zeta(t)\zeta(t') \rangle_{t' < t} = \lambda(t)\lambda(t') \frac{d}{dt'} \left(\frac{1}{\lambda(t')} \frac{d}{dt} \left[\frac{1}{\lambda(t)} \int_0^{t'} ds \lambda(t-s)\lambda(t'-s) \right] \right). \end{aligned} \quad (4.9)$$

The Langevin equation (4.7) was used in ref. [31] to construct a different set of decoupled FP, namely,

$$\begin{aligned} \frac{\partial P_1}{\partial t}(x, t) &= \frac{\partial}{\partial x} \left[A(t)x - B(t) + D(t) \frac{\partial}{\partial x} \right] P_1(x, t), \\ \frac{\partial P_2}{\partial t}(x, t, x', t') &= \frac{\partial}{\partial x} \left[A(t)x - B(t) + D(t) \frac{\partial}{\partial x} + C(t', t) \frac{\partial}{\partial x'} \right] P_2(x, t, x', t'), \end{aligned} \quad (4.10)$$

with

$$D(t) = \frac{\sigma^2 \lambda^2(t)}{2} \frac{d}{dt} \int_0^t ds \frac{\lambda^2(s)}{\lambda^2(t)}, \quad (4.11)$$

$$C(t', t) = \sigma^2 \lambda(t) \frac{d}{dt} \int_0^{t'} ds \frac{\lambda(t-s)\lambda(t'-s)}{\lambda(t)}. \quad (4.12)$$

While the coefficients of the FP set (4.10) do not depend on x_0 , eliminating x_0 makes the interpretation of these FP coefficients challenging. The drift term and the time-dependent diffusion constant are not anymore, respectively, the derivative of the mean and of the MSD of the original Langevin equation. In addition, the transformation (4.6) loses validity in the oscillatory regime. At specific points in time, corresponding to whenever $\lambda(t)$ crosses zero, the FP coefficient

$A(t)$ blows up, even though the solution of the FP is well-behaved because the singularities of the drift term are counterbalanced by singularities of opposite sign in the diffusion coefficient [31].

(b) Exact solutions

Solutions for all times for Eqs. (4.2) or (4.10) may be found analytically for natural boundary conditions $P_1, P_2 \rightarrow 0$ as $|x|, |x'| \rightarrow \infty$ (see Appendix C for the case of Eq. (4.2) and ref. [31] for Eq. (4.10)). The functional form of P_1 and P_2 is Gaussian and bivariate Gaussian, respectively. For localised initial preparation they are respectively

$$P_1(x, t|x_0, 0) = [2\pi\nu(t)]^{-\frac{1}{2}} \exp \left\{ -\frac{(x - x_0\lambda(t) - \Psi(t))^2}{2\nu(t)} \right\}, \quad (4.13)$$

and

$$P_2(x, t; x', t'|x_0, 0) = [4\pi^2\nu(t)\nu(t')(1-r^2)]^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2(1-r^2)} \left[\left(\frac{x - x_0\lambda(t) - \Psi(t)}{\nu(t)^{\frac{1}{2}}} \right)^2 + \left(\frac{x' - x_0\lambda(t') - \Psi(t')}{\nu(t')^{\frac{1}{2}}} \right)^2 - 2r \left(\frac{x - x_0\lambda(t) - \Psi(t)}{\nu(t)^{\frac{1}{2}}} \right) \left(\frac{x' - x_0\lambda(t') - \Psi(t')}{\nu(t')^{\frac{1}{2}}} \right) \right] \right\}. \quad (4.14)$$

The two-time correlation function

$$r(t', t) = \frac{\mathcal{C}(t', t)}{\sqrt{\nu(t)\nu(t')}}, \quad (4.15)$$

controls the degree of mixing along x and x' and becomes identically 0 at long times.

As Eq. (4.13) and (4.14) are the propagators for localised initial preparation, the solutions W_1 and W_2 for generic initial preparation $P_1(x, 0) = \mathcal{I}(x)$ are simply obtained by integrating over $\mathcal{I}(x)$

$$\begin{aligned} W_1(x, t) &= \int_{-\infty}^{+\infty} dx_0 P_1(x, t|x_0, 0) \mathcal{I}(x_0), \\ W_2(x, t; x', t') &= \int_{-\infty}^{+\infty} dx_0 P_2(x, t; x', t'|x_0, 0) \mathcal{I}(x_0), \end{aligned} \quad (4.16)$$

and the resulting one-time conditional probability is given by

$$Q_{1|1}(x, t|x', t')_{\mathcal{I}} = \frac{W_2(x, t; x', t')}{W_1(x', t')}, \quad (4.17)$$

where the subindex \mathcal{I} indicates a non-localised initial preparation.

5. Deriving probability distributions without constructing a Fokker-Planck equation

For completeness, we present in brief the characteristic functional method which allows to derive P_n from any Langevin equation of the form (3.1). The method was developed originally by Budini and Cáceres [20,21] for generic linear non-Markov Langevin equations and has been used in ref. [29] to find exact joint-time pdfs for single and multiple DLE.

A characteristic functional for a stochastic process $x(t)$ is a generalization of a characteristic function to

$$G_x([\kappa(t)]) = \left\langle \exp \left\{ i \int_0^\infty dt \kappa(t) x(t) \right\} \right\rangle, \quad (5.1)$$

where $\kappa(t) = k_0\delta(t) + \sum_{i=1}^{n-1} k_i\delta(t - \tau_i)$ represents a test function that selects appropriate time values. A convenient choice for these times are multiples of the delay time τ but other choices are also possible. Each term of the exponential expansion of Eq. (5.1) can be computed explicitly

using the correlation properties of the noise, and when the term of the series are resummed the result is

$$P_n(\mathbf{x}, t) = \frac{1}{\sqrt{(2\pi)^n |\Sigma(t)|}} \exp \left\{ -\frac{1}{2} [\mathbf{x} - \mathbf{K}(t)]^T \Sigma^{-1}(t) [\mathbf{x} - \mathbf{K}(t)] \right\}, \quad (5.2)$$

where \mathbf{x} is a column vector of the variable x_i for $i = 1, \dots, n$, that is $x(s)$ evaluated, respectively, at time $s = t_1, t_2, \dots, t_n$, respectively, with

$$\mathbf{K}(t) = \begin{pmatrix} K(t_1) \\ K(t_2) \\ \vdots \\ K(t_n) \end{pmatrix}, \quad (5.3)$$

and

$$\Sigma(t) = \begin{pmatrix} \mathcal{C}(t_1, t_1) & \mathcal{C}(t_1, t_2) & \dots & \mathcal{C}(t_1, t_n) \\ \mathcal{C}(t_1, t_2) & \mathcal{C}(t_2, t_2) & \dots & \mathcal{C}(t_2, t_n) \\ \dots & \dots & \ddots & \dots \\ \mathcal{C}(t_n, t_1) & \mathcal{C}(t_n, t_2) & \dots & \mathcal{C}(t_n, t_n) \end{pmatrix}. \quad (5.4)$$

In Eq. (5.3), $K(t)$ is given by Eq. (3.3) and the elements of the covariance matrix (5.4) are defined in Eq. (4.3). The functional (5.1) is a multivariate Gaussian distribution in the Fourier domain, whose inverse Fourier transform is the full time-dependent probability distribution

6. Distinguishing between Markov and non-Markov dynamics

Knowing the analytic form of the conditional distribution may allow to test from observations whether a process under investigation is Markov or non-Markov. In Sec. 2 we have pointed out that for Markov processes higher order conditional distributions can be expressed as a product of $Q_{1|1}$ conditional distributions. As we are able to construct any conditional distributions from P_n and P_{n+1} , it is instructive to determine the necessary and sufficient conditions for $Q_{1|n}$ to reduce to the product of $Q_{1|1}$. Given the cumbersome nature of the expressions, however, we limit ourselves to the case $n = 2$.

In particular, we ask when $Q_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1)_{x_0}$ loses its dependence on all past events but the most recent one. In other words we are interested in the conditions for which

$$Q_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1)_{x_0} = Q_{1|1}(x_3, t_3 | x_2, t_2), \quad (6.1)$$

where $0 < t_1 < t_2 < t_3$, and where the subscript x_0 on the left hand side indicates an explicit dependence on the initial preparation localised at $x = x_0$. This explicit dependence is no longer present on the right hand side of (6.1), and this will correspond to $Q_{1|1}$ being independent of x_0 .

We consider a generic linear Langevin equation with additive Gaussian noise and calculate $Q_{1|2}$ using the method of characteristic functionals (see Appendix D). As the Markov/non-Markov nature of the process is embedded in the form of the Green's function $\lambda(t)$, we determine the form of $\lambda(t)$ that leads to the identity (6.1).

In compact form one can write

$$Q_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1)_{x_0} = \frac{\exp \left\{ -\frac{\left(X_3(x_0, t_3) - X_2(x_0, t_2) \sqrt{\frac{\nu(t_3)}{\nu(t_2)}} \frac{[r_{23} - r_{12} r_{13}]}{[1 - r_{12}^2]} - X_1(x_0, t_1) \sqrt{\frac{\nu(t_3)}{\nu(t_1)}} \frac{[r_{13} - r_{12} r_{23}]}{[1 - r_{12}^2]} \right)^2}{2\nu(t_3)R(t_1, t_2, t_3)(1 - r_{12}^2)^{-1}} \right\}}{\left[2\pi\nu(t_3)R(t_1, t_2, t_3)(1 - r_{12}^2)^{-1} \right]^{\frac{1}{2}}}, \quad (6.2)$$

where

$$\begin{aligned} X_i &= x_i - x_0 \lambda(t_i) - \Psi(t_i), \\ r_{ij} &= \frac{\mathcal{C}(t_i, t_j)}{\sqrt{\nu(t_i)\nu(t_j)}}, \\ R(t_1, t_2, t_3) &= 1 - r_{12}^2 - r_{23}^2 - r_{13}^2 + 2r_{12}r_{23}r_{13}. \end{aligned} \quad (6.3)$$

The ratio of P_2 and P_1 gives instead

$$Q_{1|1}(x_3, t_3 | x_2, t_2)_{x_0} = \frac{\exp \left\{ - \frac{\left(X_3(x_0, t_3) - X_2(x_0, t_2) \sqrt{\frac{\nu(t_3)}{\nu(t_2)}} r_{23} \right)^2}{2\nu(t_3)(1-r_{23})} \right\}}{[2\pi\nu(t_3)(1-r_{23}^2)]^{\frac{1}{2}}}. \quad (6.4)$$

Comparison of $Q_{1|2}$ and $Q_{1|1}$ shows that Eq. (6.1) is satisfied when (6.2) and (6.4) have no explicit dependence on x_1 and x_0 , which occurs when

$$r_{13} = r_{12}r_{23}, \quad (6.5)$$

and

$$X_3(x_0, t_3) - X_2(x_0, t_2) \sqrt{\frac{\nu(t_3)}{\nu(t_2)}} r_{23} = f(t_2, t_3), \quad (6.6)$$

where f is some arbitrary function independent of x_0 .

We show first the necessary and sufficient conditions for Eq. (6.5) to be valid. We start by assuming $\lambda(t) = e^{-At}$. Using the definition of the two-time correlation function $r_{ij} = \mathcal{C}(t_i, t_j) / \sqrt{\nu(t_i)\nu(t_j)}$, and setting $\lambda(t) = e^{-At}$, it is possible to write r_{ij} in the form

$$r_{ij} = \frac{(e^{2At_i} - 1)^{\frac{1}{2}}}{(e^{2At_j} - 1)^{\frac{1}{2}}}, \quad (6.7)$$

from which (6.5) follows immediately.

To show the inverse, we rewrite Eq. (6.5) as $\mathcal{C}(t_1, t_3)\nu(t_2) = \mathcal{C}(t_1, t_2)\mathcal{C}(t_2, t_3)$. Using the definitions (3.4) and (4.3), this can be expressed explicitly in terms of the Green's function $\lambda(t)$ as

$$\int_0^{t_1} \int_0^{t_2} \lambda(t_1 - s) \lambda(t_3 - s) \lambda^2(t_2 - v) dv ds = \int_0^{t_1} \int_0^{t_2} \lambda(t_1 - s) \lambda(t_2 - s) \lambda(t_2 - v) \lambda(t_3 - v) dv ds. \quad (6.8)$$

Since no particular form was assumed for $\lambda(t)$, it follows that the integrands on both sides of Eq. (6.8) must be equal for every $t_1 < t_2 < t_3$ and for each $s \in [0, t_1]$ and $v \in [0, t_2]$. Hence, fixing values for s and v in their respective ranges, it follows that

$$\frac{\lambda(t_3 - s)}{\lambda(t_2 - s)} = \frac{\lambda(t_3 - v)}{\lambda(t_2 - v)}. \quad (6.9)$$

Since this holds for arbitrary values of s and v , where in general it is possible for $s \neq v$, it follows that both sides are independent of s and v . This is only possible if the Green's function has the form of an exponential $\lambda(t) = e^{-At}$, for a constant $A \geq 0$. The case $A = 0$ corresponds to the Wiener process, while $A > 0$ corresponds to an Ornstein-Uhlenbeck process.

Using identity (6.5) and the definitions of $\nu(t)$ and $r(t', t)$, and after substitution of the known form for X_2 and X_3 , condition (6.6) implies

$$x_0 \lambda(t_3) - x_0 \lambda(t_2) \sqrt{\frac{\nu(t_3)}{\nu(t_2)}} r_{23} = 0, \quad (6.10)$$

whose necessary and sufficient condition is once again that $\lambda(t) = e^{-At}$ for a constant $A \geq 0$.

As the necessary and sufficient conditions for a process to be Markov is a Green's function of the form $\lambda(t) = e^{-At}$, fitting of spatio-temporal data to $Q_{1|1}(x, t|x', t')$ may allow to discern the existence of underlying non-Markovian features. A clear signature of a non-Markov system compared to a Markov system is the phenomenon of ageing, for which the system is not time-translational invariant and the initial preparation of the system affects the future dynamics.

In Fig. 1 we have indicated pictorially how, in a Markov system, the future dynamics are defined by the span of time from the initial to the final observation, irrespective of the system's initial preparation. On the other hand, this is not the case in a non-Markov system. With the analytic expressions for the conditional pdf, this intuitive explanation can now be made rigorous by studying $Q_{1|1}(x, t|x', t')$ in Eq. (4.17). Except for the Markov case, one can show that the initial preparation $\mathcal{I}(x)$ does not cancel out between the numerator and denominator in the expression for $Q_{1|1}(x, t|x', t')_{\mathcal{I}}$ (see ref. [31]).

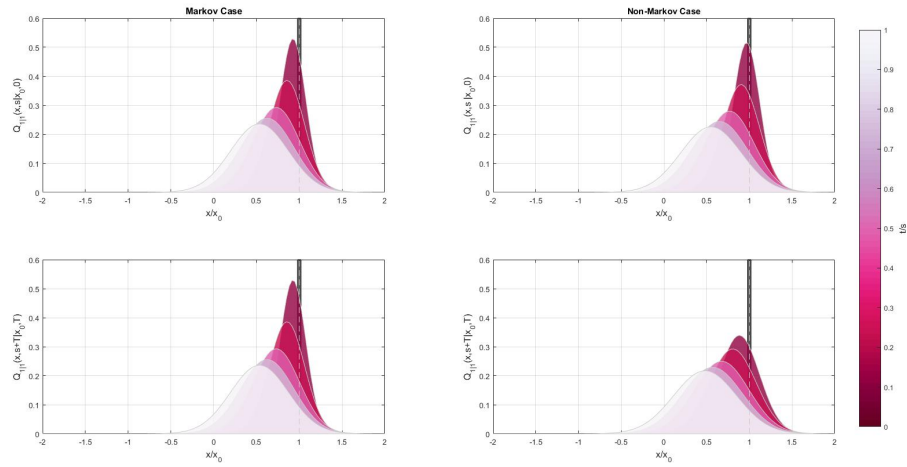


Figure 3. Evidence of ageing effects in non-Markov compared to Markov systems. The top panels display the conditional pdfs $Q_{1|1}(x, s|x_0, 0)$, while the bottom panels display the conditional pdf $Q_{1|1}(x, s+T|x_0, T)$ at time $t' = T$ (arb. units) and $x' = x_0$. The Wiener process, specifically Eq. (2.1) with $\gamma = 0$, is chosen as a Markov case (left panels). The DLE case (6.4) is the chosen non-Markov system (right panels). The values of $\sigma^2 s$ and $\sigma^2/(2\gamma)$ are set equal to 1 in arbitrary units, respectively, in the diffusive and DLE case. To avoid observing oscillations of the mean of the pdf the value $\gamma\tau = 0.3$ is chosen so as to display the characteristic dynamics of the DLE in the monotonic regime ($\gamma\tau < e^{-1}$). The vertical bar around $x = x_0$ represents the Dirac delta localisation at time $t' = 0$ and $t' = T$. The colour bar represents the time from $s = 0$ to $s = 1$ in arbitrary units.

The lack of time-translational invariance in non-Markov systems is best shown in a graphical comparison. We have done so in Fig. 3 by plotting the dynamics of $Q_{1|1}(x, s + t'|x_0, t')$ for a simple diffusive scenario in the left panels and for the DLE (3.1) in the right panels. In the top panels the conditional pdfs are shown from $t' = 0$, that is from the initial preparation localised at $x = x_0$, over a time span from $s = 0$ to $s = 1$ (arb. units), whereas in the bottom panels they are shown over the same time span, and also when localised at $x = x_0$ but starting from time $t' = T$. While the conditional pdfs are identical in the diffusive case, a difference in the spatio-temporal dynamics is visible in the delay case. As time progresses, this difference diminishes and eventually disappears when the system reaches the steady state given by $Q_{1|1}(x, \infty|x', t') = \exp[-x^2/2\nu(\infty)]/[2\pi\nu(\infty)]^{1/2}$ with $\nu(\infty) = \sigma^2[1 + \sin(\gamma\tau)]/[2\gamma \cos(\gamma\tau)]$.

7. Conclusion

In recent years, there has been renewed interest to develop a mathematical description for non-Markov systems, since a great deal of theoretical and empirical literature points to the important role that delays and memory play in many natural processes. Delay dynamics have been applied in climate models [34], and across scales in biology: from gene regulatory networks at the single cell level [35,36], to organism-scale neural networks [37], and even multi-organism collective movement [38], consensus [39] and population dynamics [40].

A clear focus to study non-Markov processes is present in biological neural networks, where communication delays can be caused by the time it takes for signals to propagate down different axon lengths. These delayed dynamics modulate the behaviour of individual neurons to produce a myriad of observable outputs, through complex collective firing patterns [41–43]. Current efforts in the field of machine learning focus on understanding spiking neural networks, in which learning dynamics are dictated by time lags between input and output activity spikes of connected neurons [44,45]. Formal developments in the mathematics used by such theories should lead to greater understanding of how this class of machine learning algorithms should operate.

Delays caused by physical sensory limitations are found in organismal biology, where delayed reactions to changes in movement direction have been shown to play an important role in coordinated flight in pigeons [46] and bats [38]. Non-Markov processes do not occur only due to sensory limitations, they may also correspond to an individual using memory capabilities and having direct access to information from the past. For example, individual animals can keep track of locations previously inhabited by themselves or others during foraging or migration. This can be done either through high cognitive abilities [47] or through stigmergy, when collective memory is created through deposition of information in the environment [48–50]. This memory will influence the choices made by individuals over which locations to explore in the future.

We have presented an overview of the stage of development of the mathematical formalisms for non-Markov process. We have discussed what has been done in the past, reviewed the most recent studies on the equivalence between the DLE and probabilistic equations, and FP equations in particular. We have presented a new FP equation to represent the DLE and compared it to an earlier form. We have also introduced an infinite hierarchy of FP equations to represent the more general problem of non-linear delayed Langevin equation with additive Gaussian noise.

While the paper has focused on delayed systems, the formalism used for the DLE can be applied to the GLE, a Langevin equation of the type

$$\dot{x}(t) = -\gamma \int_0^t x(s)\phi(t-s)ds + \sigma\xi(t), \quad x(0) = x_0, \quad (7.1)$$

which is a special form of distributed delays (see e.g. [51]). Furthermore, the general FP hierarchy presented in Sec. 2, can be employed for a non-linear version of Eq. (7.1) by replacing the delayed variable $x(t - \tau)$ with the integrated variable $\int_0^t ds\phi(t-s)x(s)$.

We hope that the generality of the approach presented in this article will enhance the burgeoning interest in non-Markov systems and help the community build a general framework to construct FP equation for non-linear delayed and distributed delays processes.

Appendix A: construction of the infinite FP hierarchy

Two steps are necessary to obtain the hierarchically FPE in Sec. 2(b). First one needs to realise that

$$\begin{aligned} & \langle f(x(t_n), x(t_n - \tau)) \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \rangle \\ &= \left\langle \int dx' f(x(t_n), x') \delta(x_n - x(t_n)) \delta(x' - x(t')) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle \\ &= \int dx' f(x_n, x') P_{n+1}(x_n, t_n; x', t'; x_{n-1}, t_{n-1}; \dots, x_1, t_1) \end{aligned} \quad (\text{A2})$$

where $t' = t_n - \tau$ and $x' = x(t_n - \tau)$ and it is assumed that t' is the second largest time after t_n . One also needs to calculate, with the help of Novikov's theorem on averages of zero of means Gaussian functional [52], the following

$$\begin{aligned} & \langle \xi(t_n) \delta(x_1 - x(t_1)) \delta(x_2 - x(t_2)) \dots \delta(x_n - x(t_n)) \rangle \\ &= \int_0^{t_n} ds \langle \xi(t_n) \xi(s) \rangle \left\langle \frac{\delta[\delta(x_n - x(t_n))]}{\delta \xi(s)} \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle \\ &+ \int_0^{t_{n-1}} ds' \langle \xi(t_n) \xi(s') \rangle \left\langle \delta(x_n - x(t_n)) \frac{\delta[\delta(x_{n-1} - x(t_{n-1}))]}{\delta \xi(s')} \dots \delta(x_1 - x(t_1)) \right\rangle + \dots \\ &+ \int_0^{t_1} ds' \langle \xi(t_n) \xi(s') \rangle \left\langle \delta(x_n - x(t_n)) \dots \delta(x_2 - x(t_2)) \frac{\delta[\delta(x_1 - x(t_1))]}{\delta \xi(s')} \right\rangle \\ &= - \int_0^{t_n} ds \langle \xi(t_n) \xi(s) \rangle \frac{\partial}{\partial x_n} \left\langle \left[\frac{\delta x(t_n)}{\delta \xi(s)} \right]_{x(t_n)=x_n} \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle \\ &- \int_0^{t_{n-1}} ds' \langle \xi(t_n) \xi(s') \rangle \frac{\partial}{\partial x_{n-1}} \left\langle \left[\frac{\delta x(t_{n-1})}{\delta \xi(s')} \right]_{x(t_{n-1})=x_{n-1}} \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \right. \\ &\quad \left. \times \dots \delta(x_1 - x(t_1)) \right\rangle + \dots \\ &- \int_0^{t_1} ds' \langle \xi(t_n) \xi(s') \rangle \frac{\partial}{\partial x_1} \left\langle \left[\frac{\delta x(t_1)}{\delta \xi(s')} \right]_{x(t_1)=x_1} \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle \\ &= - \sum_{i=1}^n \frac{\partial}{\partial x_i} \int_0^{t_i} ds \langle \xi(t_n) \xi(s) \rangle \left[\frac{\delta x(t_i)}{\delta \xi(s)} \right]_{x(t_i)=x_i} \left\langle \delta(x_n - x(t_n)) \delta(x_{n-1} - x(t_{n-1})) \dots \delta(x_1 - x(t_1)) \right\rangle \end{aligned} \quad (\text{A3})$$

The calculation proceeds depending on the type of noise. For white noise $\langle \xi(t_n) \xi(s) \rangle = \delta(t_n - s)$, and with $\frac{\delta x(t_i)}{\delta \xi(s)} = \sigma$ (see the end of Appendix B), we obtain $\int_0^{t_i} ds \langle \xi(t_n) \xi(s) \rangle = \delta_{n,i}/2$, since $t_n \geq t_i$ for all i in the sum.. Recalling that $\sigma^2/2 = D$ Eq. (A3) results in Eq. (2.10) in the main text.

Appendix B: one and two-time FP for a linear DLE

When the noise is not white in Eq. (A3), the two-time correlation $\langle \xi(t_n)\xi(s) \rangle$ is no longer a Dirac delta. To derive the coupled Eq. (4.2) one takes Eq. (4.1) and writes

$$\frac{\partial P_1}{\partial t} = -\frac{\partial}{\partial x} \left\langle \left(\dot{K}(t) + \sigma \dot{\mu}(t) \right) \delta(x - x(t)) \right\rangle, \quad (\text{B2})$$

$$\frac{\partial P_2}{\partial t} = -\frac{\partial}{\partial x} \left\langle \left(\dot{K}(t) + \sigma \dot{\mu}(t) \right) \delta(x - x(t)) \delta(x' - x(t')) \right\rangle. \quad (\text{B3})$$

It then follows that

$$\frac{\partial P_1}{\partial t} = -\frac{\partial}{\partial x} \left[\dot{K}(t) P_1 \right] - \frac{\partial}{\partial x} \langle \sigma \dot{\mu}(t) \delta(x - x(t)) \rangle, \quad (\text{B4})$$

$$\frac{\partial P_2}{\partial t} = -\frac{\partial}{\partial x} \left[\dot{K}(t) P_2 \right] - \frac{\partial}{\partial x} \langle \sigma \dot{\mu}(t) \delta(x - x(t)) \delta(x' - x(t')) \rangle. \quad (\text{B5})$$

The evaluation of the second term in (B4) and (B5) proceeds from Eq. (A3) giving

$$\langle \dot{\mu}(t) \delta(x - x(t)) \rangle = -\frac{\partial}{\partial x} \int_0^t ds \langle \dot{\mu}(t) \dot{\mu}(s) \rangle \left\langle \delta(x - x(t)) \left[\frac{\delta x(t)}{\delta \dot{\mu}(s)} \right]_{x(t)=x} \right\rangle, \quad (\text{B6})$$

and

$$\begin{aligned} & \sigma \langle \dot{\mu}(t) \delta(x - x(t)) \delta(x' - x(t')) \rangle \\ &= \sigma \int_0^t ds \langle \dot{\mu}(t) \dot{\mu}(s) \rangle \left\langle \frac{\delta[\delta(x - x(t))]}{\delta \dot{\mu}(s)} \delta(x' - x(t')) \right\rangle \\ &+ \sigma \int_0^{t'} ds' \langle \dot{\mu}(t) \dot{\mu}(s') \rangle \left\langle \delta(x - x(t)) \frac{\delta[\delta(x' - x(t'))]}{\delta \dot{\mu}(s')} \right\rangle, \\ &= -\sigma \frac{\partial}{\partial x} \int_0^t ds \langle \dot{\mu}(t) \dot{\mu}(s) \rangle \left\langle \delta(x - x(t)) \delta(x' - x(t')) \left[\frac{\delta x(t)}{\delta \dot{\mu}(s)} \right]_{x(t)=x} \right\rangle \\ &- \sigma \frac{\partial}{\partial x} \int_0^{t'} ds' \langle \dot{\mu}(t) \dot{\mu}(s') \rangle \left\langle \delta(x - x(t)) \delta(x' - x(t')) \left[\frac{\delta x(t')}{\delta \dot{\mu}(s')} \right]_{x(t')=x'} \right\rangle, \end{aligned} \quad (\text{B7})$$

where $[\delta x(t)/\delta \dot{\mu}(s)]_{x(t)=x}$ represents the functional derivative of the Langevin equation solution (3.3) with respect to $\dot{\mu}$, evaluated at the value $x(t) = x$. To evaluate this derivative, we use a relationship analogous to the chain rule for functional derivatives of functions obeying the general Langevin equation

$$\dot{x}(t) = F(x, t) + \sigma \eta(t), \quad (\text{B9})$$

where $\eta(t)$ is a general stochastic variable, and in this example $\eta(t) = \dot{\mu}(t)$. For such a function $x(t)$

$$\begin{aligned} \frac{\partial}{\partial t} \left(\frac{\delta x(t)}{\delta \dot{\mu}(s)} \right) &= \frac{\delta x(t)}{\delta \dot{\mu}(s)} \frac{\partial F(x, t)}{\partial x}, \\ &= \frac{\delta x(t)}{\delta \dot{\mu}(s)} \frac{\partial}{\partial x} \left(\dot{K}(t) \right), \\ &= 0, \end{aligned} \quad (\text{B10})$$

where in the second line we have considered the specific form of the Langevin equation. With the solution of Eq. (B10) being

$$\frac{\delta x(t)}{\delta \dot{\mu}(s)} = \sigma, \quad (\text{B11})$$

one only needs to evaluate the remaining averages of the noise-terms

$$\begin{aligned}
 & \int_0^t ds \langle \dot{\mu}(t) \dot{\mu}(s) \rangle \\
 &= \int_0^t ds \left\langle \frac{d}{dt} \left(\int_0^t \lambda(t-v) \xi(v) dv \right) \frac{d}{ds} \left(\int_0^s \lambda(s-v') \xi(v') dv' \right) \right\rangle, \\
 &= \left\langle \frac{d}{dt} \left(\int_0^t \lambda(t-v) \xi(v) dv \right) \int_0^t \lambda(t-v') \xi(v') dv' \right\rangle, \\
 &= \frac{1}{2} \frac{d}{dt} \left(\int_0^t \int_0^t \lambda(t-v) \lambda(t-v') \langle \xi(v) \xi(v') \rangle dv dv' \right), \\
 &= \frac{1}{2} \frac{d}{dt} \left(\int_0^t \int_0^t \lambda(t-v) \lambda(t-v') \delta(v-v') dv dv' \right), \\
 &= \frac{1}{2} \frac{d}{dt} \left(\int_0^t \lambda^2(v) dv \right)
 \end{aligned} \tag{B12}$$

and

$$\begin{aligned}
 & \int_0^{t'} ds' \langle \dot{\mu}(t) \dot{\mu}(s') \rangle \\
 &= \int_0^{t'} ds' \left\langle \frac{d}{dt} \left(\int_0^t \lambda(t-v) \xi(v) dv \right) \frac{d}{ds'} \left(\int_0^{s'} \lambda(s'-v') \xi(v') dv' \right) \right\rangle, \\
 &= \left\langle \frac{d}{dt} \left(\int_0^t \lambda(t-v) \xi(v) dv \right) \int_0^{t'} \lambda(t'-v') \xi(v') dv' \right\rangle, \\
 &= \frac{d}{dt} \left(\int_0^{t'} \int_0^t \lambda(t-v) \lambda(t'-v') \langle \xi(v) \xi(v') \rangle dv dv' \right), \\
 &= \frac{d}{dt} \left(\int_0^{t'} \int_0^t \lambda(t-v) \lambda(t'-v') \delta(v-v') dv dv' \right), \\
 &= \frac{d}{dt} \left(\int_0^{t'} \lambda(t-v') \lambda(t'-v') dv' \right).
 \end{aligned} \tag{B13}$$

In the second line of (B12), we evaluated the integral in s to change the limits of the integral in v' . In the third line we combined the two integrals into one derivative, using a form of the product rule, which introduces a factor of $1/2$ since both integrals have a t dependence. We then evaluate the average over the noise terms which produces a Dirac delta function that changes the variable v' to v inside the integrals. In the third line of (B13), the integrals were combined into the t derivative and no factor $1/2$ was introduced since only one of the terms has a t dependence. Notice also that for the Wiener process, $\lambda(t) = 1$ and $\int_0^t ds \langle \dot{\mu}(t) \dot{\mu}(s) \rangle = 1/2$.

Substituting these expressions into equations (B4) and (B5) we obtain the P_1 and P_2 FP equations given by (4.2).

Appendix C

In this section, solutions of the FP (4.2) are derived. The solution method is similar for the two equations. The trick is to transform the P_1 and P_2 equations into the diffusion equation. The P_2 equation must be Fourier transformed first in one of the variables before it can be written as a diffusion equation. The solution is then inverted back to the original variables. We start with the

P_1 equation, given in Eq. (4.2), which can be transformed into the diffusion equation

$$\frac{\partial V}{\partial \rho}(y, \rho) = \frac{\partial^2 V}{\partial y^2}(y, \rho), \quad V(y, \rho = 0) = \delta(y), \quad (C2)$$

by making the variable transformation

$$P_1(x, t) = V(y, \rho), \quad y(x, t) = x - K(t), \quad \rho(t) = \frac{\sigma^2}{2} \int_0^t \lambda^2(s) ds. \quad (C3)$$

Equation (C2) has a well known solution [53] given by

$$V(y, \rho) = \frac{1}{\sqrt{4\pi\rho}} \int_{-\infty}^{\infty} V(z, 0) \exp\left\{-\frac{(y-z)^2}{4\rho}\right\} dz. \quad (C4)$$

Evaluating the initial condition, and inverting back to the original variables yields the P_1 solution (4.13).

The P_2 equation is given by

$$\frac{\partial P_2}{\partial t} = -\dot{K}(t) \frac{\partial P_2}{\partial x} + \frac{\dot{\nu}(t)}{2} \frac{\partial^2 P_2}{\partial x^2} + \frac{\partial \mathcal{C}(t', t)}{\partial t} \frac{\partial^2 P_2}{\partial x \partial x'}, \quad (C5)$$

$$P_2(x, t = t'; x', t' | x_0, 0) = \delta(x - x') P_1(x', t' | x_0, 0). \quad (C6)$$

The Fourier Transform between two variables x and k is defined here as

$$\mathcal{F}_{x \rightarrow k}[f(x)] = \int_{-\infty}^{\infty} e^{-ikx} f(x) dx, \quad (C7)$$

with inverse given by

$$\mathcal{F}_{k \rightarrow x}^{-1}[\tilde{f}(k)] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ikx} \tilde{f}(k) dk, \quad (C8)$$

Performing a Fourier Transform in $x' \rightarrow k'$ and letting $\tilde{P}_2 = \mathcal{F}_{x' \rightarrow k'}[P_2]$, the P_2 equation becomes

$$\frac{\partial \tilde{P}_2}{\partial t} = -\dot{K}(t) \frac{\partial \tilde{P}_2}{\partial x} + \frac{\dot{\nu}(t)}{2} \frac{\partial^2 \tilde{P}_2}{\partial x^2} + ik' \frac{\partial \mathcal{C}(t', t)}{\partial t} \frac{\partial \tilde{P}_2}{\partial x}, \quad (C9)$$

$$\tilde{P}_2(x, t = t'; k', t' | x_0, 0) = e^{-ik'x} P_1(x, t' | x_0, 0). \quad (C10)$$

Eq. (C9) can now be transformed into the diffusion equation (C2) using the variable transformation

$$\tilde{P}_2(x, t; k', t') = V(y, \rho), \quad y(x, t; k', t') = x - K(t) + ik' \mathcal{C}(t', t), \quad \rho(t', t) = \frac{\sigma^2}{2} \int_{t'}^t \lambda^2(s) ds. \quad (C11)$$

The initial condition is found by setting $t = t'$ in the definition of V and changing to the new variables in the initial condition in (C9). This yields

$$V(y, 0) = \exp\{-ik'(y + K(t') - ik'\nu(t'))\} P_1(y + K(t') - ik'\nu(t'), t'). \quad (C12)$$

Inserting this initial condition into the general solution (C4) and inverting to the original variables by noting that $\rho(t', t) = (\nu(t) - \nu(t'))/2$, yields

$$\begin{aligned} \tilde{P}_2(x, t; k', t') &= [4\pi\rho]^{-\frac{1}{2}} \int_{-\infty}^{\infty} dz \exp \left\{ -ik' [z + K(t') - ik'\nu(t')] \right\} \\ &\times P_1 \left(z + K(t') - ik'\nu(t'), t' \right) \exp \left\{ -\frac{[x - K(t) + ik'\mathcal{C}(t', t) - z]^2}{2[\nu(t) - \nu(t')]} \right\}. \end{aligned} \quad (\text{C13})$$

At this stage it is simpler to evaluate the integral in z by first performing a Fourier Transform in $x \rightarrow k$, and then performing a Fourier inversion in both variables.

Noting that the Fourier Transform of a shifted Gaussian is given by

$$\mathcal{F} \left[[4\pi\alpha]^{-\frac{1}{2}} e^{-\frac{(x-\beta)^2}{4\alpha}} \right] = e^{-\alpha k^2} e^{-ik\beta}, \quad (\text{C14})$$

the Fourier Transform in x yields

$$\begin{aligned} \tilde{P}_2(k, t; k', t') &= [4\pi\rho]^{-\frac{1}{2}} \int_{-\infty}^{\infty} dz \exp \left\{ -ik' [z + K(t') - ik'\nu(t')] \right\} \\ &\times P_1 \left(z + K(t') - ik'\nu(t'), t' \right) \exp \left\{ -\frac{k^2}{2}(\nu(t) - \nu(t')) - ik(z + K(t) - ik'\mathcal{C}(t', t)) \right\}. \end{aligned} \quad (\text{C15})$$

By separating out terms proportional to z , the integral may be evaluated to give

$$\tilde{P}_2(k, t; k', t') = \exp \left\{ -ik'K(t) - ikK(t) - \frac{k^2}{2}\nu(t) - \frac{k'^2}{2}\nu(t') - kk'\mathcal{C}(t', t) \right\}. \quad (\text{C16})$$

Finally, performing a Fourier inversion in k and k' yields the P_2 solution given in (4.14).

Appendix D: the conditional probability distribution $Q_{1|2}$

For three consecutive times $t_1 < t_2 < t_3$, and three consecutive positions $x(t_1) = x_1$, $x(t_2) = x_2$, $x(t_3) = x_3$, in Eq. (5.2) we have $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{K}(t) = (K(t_1), K(t_2), K(t_3))^T$ where T represents the transpose of the given vector and

$$\Sigma(t) = \begin{pmatrix} \mathcal{C}(t_1, t_1) & \mathcal{C}(t_1, t_2) & \mathcal{C}(t_1, t_3) \\ \mathcal{C}(t_1, t_2) & \mathcal{C}(t_2, t_2) & \mathcal{C}(t_2, t_3) \\ \mathcal{C}(t_1, t_3) & \mathcal{C}(t_2, t_3) & \mathcal{C}(t_3, t_3) \end{pmatrix}, \quad (\text{D2})$$

with $\mathcal{C}(t_i, t_j)$ defined in Eq. (4.3). To simplify notation we let $\mathcal{C}(t_i, t_j) = \mathcal{C}_{ij}$, and note that $\mathcal{C}_{ii} = \nu(t_i) = \nu_i$, so that

$$\Sigma(t) = \begin{pmatrix} \nu_1 & \mathcal{C}_{12} & \mathcal{C}_{13} \\ \mathcal{C}_{12} & \nu_2 & \mathcal{C}_{23} \\ \mathcal{C}_{13} & \mathcal{C}_{23} & \nu_3 \end{pmatrix}. \quad (\text{D3})$$

The determinant and inverse of the covariance matrix are respectively

$$\begin{aligned} |\Sigma(t)| &= \nu_1\nu_2\nu_3 - \nu_1\mathcal{C}_{23}^2 - \nu_2\mathcal{C}_{13}^2 - \nu_3\mathcal{C}_{12}^2 + 2\mathcal{C}_{12}\mathcal{C}_{23}\mathcal{C}_{13}, \\ &= \nu_1\nu_2\nu_3(1 - r_{12}^2 - r_{23}^2 - r_{13}^2 + 2r_{12}r_{23}r_{13}), \end{aligned} \quad (\text{D4})$$

and

$$\Sigma^{-1}(t) = \frac{1}{|\Sigma(t)|} \begin{pmatrix} \nu_2\nu_3 - \mathcal{C}_{23}^2 & \mathcal{C}_{13}\mathcal{C}_{23} - \nu_3\mathcal{C}_{12} & \mathcal{C}_{12}\mathcal{C}_{23} - \nu_2\mathcal{C}_{13} \\ \mathcal{C}_{13}\mathcal{C}_{23} - \nu_3\mathcal{C}_{12} & \nu_1\nu_3 - \mathcal{C}_{13}^2 & \mathcal{C}_{12}\mathcal{C}_{13} - \nu_1\mathcal{C}_{23} \\ \mathcal{C}_{12}\mathcal{C}_{23} - \nu_2\mathcal{C}_{13} & \mathcal{C}_{12}\mathcal{C}_{13} - \nu_1\mathcal{C}_{23} & \nu_1\nu_2 - \mathcal{C}_{12}^2 \end{pmatrix}, \quad (\text{D5})$$

where

$$r_{ij} = \frac{C_{ij}}{\sqrt{\nu_i \nu_j}}, \quad i, j = 1, 2, 3. \quad (\text{D6})$$

The exponent in Eq. (5.2) is thus

$$\begin{aligned} & [\mathbf{x} - \mathbf{K}(t)]^T \Sigma^{-1}(t) [\mathbf{x} - \mathbf{K}(t)] \\ &= \frac{\nu_1 \nu_2 \nu_3}{|\Sigma(t)|} \left\{ (1 - r_{23}^2) \frac{X_1^2}{\nu_1} + (1 - r_{13}^2) \frac{X_2^2}{\nu_2} + (1 - r_{12}^2) \frac{X_3^2}{\nu_3} \right. \\ & \quad \left. - 2 \left((r_{12} - r_{13} r_{23}) \frac{X_1 X_2}{\sqrt{\nu_1 \nu_2}} + (r_{23} - r_{12} r_{13}) \frac{X_2 X_3}{\sqrt{\nu_2 \nu_3}} + (r_{13} - r_{12} r_{23}) \frac{X_1 X_3}{\sqrt{\nu_1 \nu_3}} \right) \right\}. \end{aligned} \quad (\text{D7})$$

From the above equation, the three-time probability density function can be written down as

$$\begin{aligned} P_3(x_1, t_1; x_2, t_2, x_3, t_3 | x_0, 0) &= [8\pi^3 \nu_1 \nu_2 \nu_3 (1 - r_{12}^2 - r_{23}^2 - r_{13}^2 + 2r_{12} r_{23} r_{13})]^{-\frac{1}{2}} \\ & \exp \left\{ - \frac{1}{2(1 - r_{12}^2 - r_{23}^2 - r_{13}^2 + 2r_{12} r_{23} r_{13})} \left[(1 - r_{23}^2) \frac{X_1^2}{\nu_1} + (1 - r_{13}^2) \frac{X_2^2}{\nu_2} \right. \right. \\ & \quad \left. \left. + (1 - r_{12}^2) \frac{X_3^2}{\nu_3} - 2 \left((r_{12} - r_{13} r_{23}) \frac{X_1 X_2}{\sqrt{\nu_1 \nu_2}} + (r_{23} - r_{12} r_{13}) \frac{X_2 X_3}{\sqrt{\nu_2 \nu_3}} + (r_{13} - r_{12} r_{23}) \frac{X_1 X_3}{\sqrt{\nu_1 \nu_3}} \right) \right] \right\}. \end{aligned} \quad (\text{D8})$$

With the definition for $R(t_1, t_2, t_3)$ in Eq. (6.3), $Q_{1|2}$ may be written in the more compact form

$$\begin{aligned} Q_{1|2}(x_3, t_3 | x_2, t_2; x_1, t_1)_{x_0} &= \left[2\pi \nu(t_3) \frac{R(t_1, t_2, t_3)}{(1 - r_{12}^2)} \right]^{-\frac{1}{2}} \\ & \exp \left\{ - \frac{1}{2R(t_1, t_2, t_3)} \left[\left(\left(1 - r_{23}^2 \right) - \frac{R(t_1, t_2, t_3)}{(1 - r_{12}^2)} \right) \frac{X_1^2}{\nu(t_1)} \right. \right. \\ & \quad \left. \left. + \left(\left(1 - r_{13}^2 \right) - \frac{R(t_1, t_2, t_3)}{(1 - r_{12}^2)} \right) \frac{X_2^2}{\nu(t_2)} + \left(1 - r_{12}^2 \right) \frac{X_3^2}{\nu(t_3)} \right. \right. \\ & \quad \left. \left. - 2 \left(r_{12} - r_{13} r_{23} - \frac{r_{12} R(t_1, t_2, t_3)}{(1 - r_{12}^2)} \right) \frac{X_1 X_2}{(\nu(t_1) \nu(t_2))^{\frac{1}{2}}} \right. \right. \\ & \quad \left. \left. - 2 \left(r_{23} - r_{12} r_{13} \right) \frac{X_2 X_3}{(\nu(t_2) \nu(t_3))^{\frac{1}{2}}} - 2 \left(r_{13} - r_{12} r_{23} \right) \frac{X_1 X_3}{(\nu(t_1) \nu(t_3))^{\frac{1}{2}}} \right] \right\}. \end{aligned} \quad (\text{D9})$$

Eq. (D9) can be simplified further by noticing that the terms in the exponent may be factorised, and one obtains Eq. (6.2) of the main text.

Authors' Contributions. LG conceived and designed the study, and drafted the manuscript with input from ZN for certain sections including figures. All authors gave final approval for publication.

Competing Interests. The authors declare that they have no competing interests.

Funding. LG and ZN acknowledges funding from, respectively, the Engineering and Physical Research Council grants EP/I013717/1 and N509619/1.

Acknowledgements. LG thanks the organizers of the Delay Complex Systems workshop held at the Physikzentrum Bad Honnef, Germany, on 2-5 July 2018 during which some of the ideas for the paper originated. LG acknowledges discussions with André Longtin, Michael C. Mackey, and Toru Ohira.

References

1. Paul Langevin.
Sur la théorie du mouvement brownien.
Compt. Rendus, 146:530–533, 1908.
2. M. von Smoluchowski.

- Über Brownsche molekularbewegung unter einwirkung äußerer kräfte und deren zusammenhang mit der verallgemeinerten diffusionsgleichung.
Ann. Phys., 353(24):1103–1112, 1916.
3. H. Risken.
The Fokker-Planck equation: methods of solution and applications.
Springer Series in Synergetics, vol. 18, Berlin, Germany, 1984.
 4. R. Bellman and K. L. Cooke.
Differential-difference equations.
Rand Corporation, 1963.
 5. U. Küchler and B. Mensch.
Langevins stochastic differential equation extended by a time-delayed term.
Stoc. Stoc. Rep., 40(1):23–42, August 1992.
 6. R. F. Fox.
Analysis of nonstationary, Gaussian and non-Gaussian, generalized Langevin equations using methods of multiplicative stochastic processes.
J. Stat. Phys., 16(3):259–279, 1977.
 7. R. F. Fox.
The generalized Langevin equation with Gaussian fluctuations.
J. Math. Phys., 18(12):2331–2335, 1977.
 8. R. F. Fox.
Gaussian stochastic processes in physics.
Phy. Rep., 48(3):179–283, 1978.
 9. P. Hänggi and H. Thomas.
Time evolution, correlations, and linear response of non-Markov processes.
Z. Phys., 26(1):85–92, 1977.
 10. P. Hänggi.
Correlation functions and master equations of generalized (non-Markovian) Langevin equations.
Z. Phys., 31(4):407–416, 1978.
 11. P. Hänggi, H. Thomas, H. Grabert, and P. Talkner.
Note on time evolution of non-Markov processes.
J. Stat. Phys., 18(2):155–159, 1978.
 12. P. Hänggi and P. Talkner.
Non-Markov processes: The problem of the mean first passage time.
Z. Phys., 45(1):79–83, 1981.
 13. M. San Miguel and J. M. Sancho.
A colored-noise approach to Brownian motion in position space. Corrections to the Smoluchowski equation.
J. Stat. Phys., 22(5):605–624, 1980.
 14. J. M. Sancho, M. San Miguel, S. L. Katz, and J. D. Gunton.
Analytical and numerical studies of multiplicative noise.
Phys. Rev. A, 26(3):1589, 1982.
 15. J. M. Sancho, M. San Miguel, and D. Dürr.
Adiabatic elimination for systems of brownian particles with nonconstant damping coefficients.
J. Stat. Phys., 28(2):291–305, 1982.
 16. S. A. Adelman.
Fokker-Planck equations for simple non-markovian systems.
J. Chem. Phys., 64(1):124–130, 1976.
 17. Shoichi Okuyama and David W Oxtoby.
The generalized smoluchowski equation and non-markovian dynamics.
The Journal of chemical physics, 84(10):5824–5829, 1986.
 18. Steve Guillouezic, Ivan L?Heureux, and André Longtin.
Small delay approximation of stochastic delay differential equations.
Physical Review E, 59(4):3970, 1999.
 19. A. Longtin.
Stochastic delay-differential equations.
In F. M. Atay, editor, Complex Time-Delay Systems, Understanding Complex Systems, pages 177–195. Springer Berlin Heidelberg, 2010.

20. A. A. Budini and M. O. Cáceres.
The generalized Ornstein-Uhlenbeck process.
J. Phys. A: Math. Gen., 8427, 1997.
21. Adrián A Budini and M.O. Cáceres.
The generalized Wiener process ii: Finite systems.
J. Phys. A: Math. Gen., 32(22):4005, 1999.
22. A. A. Budini and M. O. Cáceres.
Functional characterization of generalized Langevin equations.
J. Phys. A: Math. Gen., 37(23):5959–5981, Jun 2004.
23. T. Ohira and J. G. Milton.
Delayed random walks.
Phys. Rev. E, 52(3):3277, 1995.
24. Toru Ohira and Toshiyuki Yamane.
Delayed stochastic systems.
Physical Review E, 61(2):1247, 2000.
25. T. D. Frank and P. J. Beek.
Stationary solutions of linear stochastic delay differential equations: Applications to biological systems.
Phys. Rev. E, 64(2):021917, 2001.
26. T. D. Frank.
Multivariate Markov processes for stochastic systems with delays: application to the stochastic Gompertz model with delay.
Phys. Rev. E, 66(1):011914, 2002.
27. T. D. Frank.
A Markov approach to nonlinear multivariate delay systems with noise.
Phys. Scr., 68(5):333, 2003.
28. TD Frank.
Delay Fokker-Planck equations, Novikov’s theorem, and Boltzmann distributions as small delay approximations.
Physical Review E, 72(1):011112, 2005.
29. T. J. McKetterick and L. Giuggioli.
Exact dynamics of stochastic linear delayed systems: application to spatio-temporal coordination of co-moving agents.
University of Bristol preprint, 2014.
30. M. Chase, K. Spendier, and V. M. Kenkre.
Analysis of confined random walkers with applications to processes occurring in molecular aggregates and immunological systems.
J. Phys. Chem. B, 120(12):3072–3080, 2016.
31. Luca Giuggioli, Thomas John McKetterick, VM Kenkre, and Matthew Chase.
Fokker-planck description for a linear delayed langevin equation with additive gaussian noise.
Journal of Physics A: Mathematical and Theoretical, 49(38):384002, 2016.
32. Bayes T.
An essay towards solving a problem in the doctrine of chances.
Phil. Trans., pages 370–418, 1763.
33. J Łuczka.
Non-markovian stochastic processes: Colored noise.
Chaos: An Interdisciplinary Journal of Nonlinear Science, 15(2):026107, 2005.
34. Andrew Keane, Bernd Krauskopf, and Claire M Postlethwaite.
Climate models with delay differential equations.
Chaos: An Interdisciplinary Journal of Nonlinear Science, 27(11):114309, 2017.
35. Dmitri Bratsun, Dmitri Volfson, Lev S Tsimring, and Jeff Hasty.
Delay-induced stochastic oscillations in gene regulation.
Proceedings of the National Academy of Sciences, 102(41):14593–14598, 2005.
36. Kiresh Parmar, Konstantin B Blyuss, Yuliya N Kyrychko, and Stephen John Hogan.
Time-delayed models of gene regulatory networks.
Computational and mathematical methods in medicine, 2015, 2015.
37. Anne Beuter, Jacques Bélair, and Christiane Labrie.

- Feedback and delays in neurological diseases: A modeling study using dynamical systems. *Bulletin of mathematical biology*, 55(3):525–541, 1993.
38. L. Giuggioli, T.J. McKetterick, and M. Holderied.
Delayed response and biosonar perception explain movement coordination in trawling bats. *PLoS Comput. Biol.*, 11(3):e1004089, 2015.
 39. Alexandre Seuret, Dimos V Dimarogonas, and Karl H Johansson.
Consensus under communication delays.
In Decision and Control, 2008. CDC 2008. 47th IEEE Conference on, pages 4922–4927. IEEE, 2008.
 40. Kondalsamy Gopalsamy.
Stability and oscillations in delay differential equations of population dynamics, volume 74. Springer Science & Business Media, 2013.
 41. Fatihcan M Atay and Axel Hutt.
Neural fields with distributed transmission speeds and long-range feedback delays.
SIAM Journal on Applied Dynamical Systems, 5(4):670–698, 2006.
 42. Philipp Hövel, Markus A Dahlem, and Eckehard Schöll.
Synchronization of noise-induced oscillations by time-delayed feedback.
In AIP Conference Proceedings, volume 922(1), pages 595–598. AIP, 2007.
 43. Markus A Dahlem, G Hiller, Anastasiia Panchuk, and Eckehard Schöll.
Dynamics of delay-coupled excitable neural systems.
International Journal of Bifurcation and Chaos, 19(02):745–753, 2009.
 44. Amirhossein Tavanaei, Masoud Ghodrati, Saeed Reza Kheradpisheh, Timothee Masquelier, and Anthony S Maida.
Deep learning in spiking neural networks.
arXiv preprint arXiv:1804.08150, 2018.
 45. Filip Ponulak and Andrzej Kasinski.
Introduction to spiking neural networks: Information processing, learning and applications.
Acta neurobiologiae experimentalis, 71(4):409–433, 2011.
 46. M. Nagy, Z. Akos, D. Biro, and T. Vicsek.
Hierarchical group dynamics in pigeon flocks.
Nature, 464:890–893, 2010.
 47. Chloe Bracis, Eliezer Gurarie, Bram Van Moorter, and R Andrew Goodwin.
Memory effects on movement behavior in animal foraging.
PloS one, 10(8):e0136057, 2015.
 48. Guy Theraulaz and Eric Bonabeau.
A brief history of stigmergy.
Artificial life, 5(2):97–116, 1999.
 49. L. Giuggioli, J. R. Potts, D. I. Rubenstein, and S. A. Levin.
Stigmergy, collective actions, and animal social spacing.
Proc. Natl. Acad. Sci. USA, 110(42):16904–16909, 2013.
 50. L. Giuggioli and V. M. Kenkre.
Consequences of animal interactions on their dynamics: emergence of home ranges and territoriality.
Mov. Ecol., 2, 2014.
 51. R Jessop and Sue Ann Campbell.
Approximating the stability region of a neural network with a general distribution of delays.
Neural Networks, 23(10):1187–1201, 2010.
 52. Evgenii A Novikov.
Functionals and the random-force method in turbulence theory.
Sov. Phys. JETP, 20(5):1290–1294, 1965.
 53. Richard Haberman.
Elementary applied partial differential equations, volume 987. Prentice Hall Englewood Cliffs, NJ, 1983.